

#2
aA
4/2/02

PATENT

Docket No. JP9-2000-0337 (291)

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Application of ITOH, et al.

Application No.

Examiner:

Filed: (Herewith)

Group Art Unit:

For: SPEECH RECOGNITION APPARATUS AND COMPUTER SYSTEM
THEREFOR, SPEECH RECOGNITION METHOD AND PROGRAM AND
RECORDING MEDIUM THEREFOR



CLAIM OF FOREIGN PRIORITY

Box Patent Application
Commissioner for Patents
Washington, D.C. 20231

Sir:

Priority under the International Convention for the Protection of Industrial
Property and under 35 U.S.C. §119 is hereby claimed for the above-identified patent
application, based upon Japanese Application No. 2001-044186 filed February 20,
2001, and a certified copy of this application is submitted herewith which perfects the
Claim of Foreign Priority.

Respectfully submitted,

Date: 1/24/02

Kevin T. Cuenot

Gregory A. Nelson, Registration No. 30,577
Kevin T. Cuenot, Registration No. 46,283
Steven M. Greenberg, Registration No. 44,725
AKERMAN SENTERFITT
222 Lakeview Avenue
Post Office Box 3188
West Palm Beach, FL 33402-3188
Telephone: (561) 653-5000

Docket No. JP9-2000-0337 (291)

Express Mailing Label No. EL 649719607 US

日 本 国 特 許 庁
JAPAN PATENT OFFICE

J1046 U.S. PTO
10/056149
01/24/02

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office

出 願 年 月 日

Date of Application:

2001年 2月20日

出 願 番 号

Application Number:

特願2001-044186

出 願 人

Applicant(s):

インターナショナル・ビジネス・マシーンズ・コーポレーション

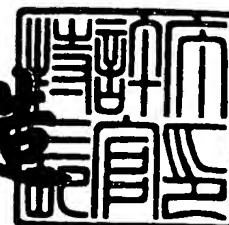
BEST AVAILABLE COPY

CERTIFIED COPY OF
PRIORITY DOCUMENT

2001年 6月21日

特許庁長官
Commissioner,
Japan Patent Office

及 川 耕 造



出証番号 出証特2001-3058839

【書類名】	特許願
【整理番号】	JP9000337
【提出日】	平成13年 2月20日
【あて先】	特許庁長官 殿
【国際特許分類】	G10L 15/00
【発明者】	
【住所又は居所】	神奈川県大和市下鶴間 1 6 2 3 番地 1 4 日本アイ・ピー・エム株式会社 東京基礎研究所内
【氏名】	伊東 伸泰
【発明者】	
【住所又は居所】	神奈川県大和市下鶴間 1 6 2 3 番地 1 4 日本アイ・ピー・エム株式会社 東京基礎研究所内
【氏名】	西村 雅史
【特許出願人】	
【識別番号】	390009531
【氏名又は名称】	インターナショナル・ビジネス・マシーンズ・コーポレーション
【代理人】	
【識別番号】	100086243
【弁理士】	
【氏名又は名称】	坂口 博
【代理人】	
【識別番号】	100091568
【弁理士】	
【氏名又は名称】	市位 嘉宏
【代理人】	
【識別番号】	100106699
【弁理士】	
【氏名又は名称】	渡部 弘道

【復代理人】

【識別番号】 100104880

【弁理士】

【氏名又は名称】 古部 次郎

【選任した復代理人】

【識別番号】 100100077

【弁理士】

【氏名又は名称】 大場 充

【手数料の表示】

【予納台帳番号】 081504

【納付金額】 21,000円

【提出物件の目録】

【物件名】 明細書 1

【物件名】 図面 1

【物件名】 要約書 1

【包括委任状番号】 9706050

【包括委任状番号】 9704733

【包括委任状番号】 0004480

【プルーフの要否】 要

【書類名】 明細書

【発明の名称】 音声認識装置、コンピュータ・システム、音声認識方法、プログラムおよび記録媒体

【特許請求の範囲】

【請求項 1】 音声に含まれる音素列を 1 つ以上の単語列に変換し、変換して得られた前記単語列それぞれに対して、前記音素列により本来示される単語列であることの可能性を示す出現確率を付与する変換手段と、

前記変換手段により得られた前記単語列それぞれに対応する言語モデルが示す更新用数値に基づいて、前記変換手段により前記単語列に付与された出現確率の更新を行う更新手段と、

更新された出現確率が、前記音素列により本来示される単語列が最も高いことを示す前記単語列のいずれかを選択し、前記音声を認識する認識手段とを備え、

前記更新手段は、自由発話特有の表現に特化した第 1 の言語モデルと、前記第 1 の言語モデルとは異なる第 2 の言語モデルとを用いて前記更新用数値を算出し、算出した前記更新用数値に基づいて、前記出現確率の更新を行うことを特徴とする音声認識装置。

【請求項 2】 前記第 1 の言語モデルは、自由発話特有の表現に含まれる所定の単語を含む単語列が、前記音素列により本来示される単語列である確からしさを示すものであることを特徴とする請求項 1 に記載の音声認識装置。

【請求項 3】 前記所定の単語を音声認識の結果に含める場合には、
前記変換手段は前記音素列を前記所定の単語を含む単語列に変換し、
前記更新手段は、前記第 1 の言語モデルおよび前記第 2 の言語モデルに基づいて、前記単語列それぞれの出現確率を更新する、
ことを特徴とする請求項 2 に記載の音声認識装置。

【請求項 4】 前記第 1 の言語モデルは、冗長語を含む単語組みを要素とすることを特徴とする請求項 1 に記載の音声認識装置。

【請求項 5】 前記第 1 の言語モデルおよび前記第 2 の言語モデルは、N-gramモデルであり、

前記更新手段は、前記第 1 の言語モデルおよび前記第 2 の言語モデルの加重平

均値を前記更新用数値として用いることを特徴とする請求項 1 に記載の音声認識装置。

【請求項 6】 請求項 1 ～ 5 のいずれかに記載の音声認識装置を有するコンピュータ・システム。

【請求項 7】 音声に含まれる音素列を 1 つ以上の単語列に変換し、変換して得られた前記単語列それぞれに対して、前記音素列により本来示される単語列であることの可能性を示す出現確率を付与する変換ステップと、

前記変換ステップにより得られた前記単語列が自由発話固有の単語を含む場合に、自由発話固有の表現に特化した第 1 の言語モデルおよび前記第 1 の言語モデルとは異なる第 2 の言語モデルを参照することにより、前記単語列それぞれに付された出現確率の更新を行う更新ステップと、

更新された出現確率が、前記音素列により本来示される単語列が最も高いことを示す前記単語列のいずれかを選択し、前記音声を認識する認識ステップと、を備えたことを特徴とする音声認識方法。

【請求項 8】 前記第 1 の言語モデルは、自由発話固有の単語を含む単語列の出現確率を、連続する N 個の単語の組み合わせと対応付けて記述するものであることを特徴とする請求項 7 に記載の音声認識方法。

【請求項 9】 前記自由発話固有の単語が、冗長語であることを特徴とする請求項 8 に記載の音声認識方法。

【請求項 10】 前記更新ステップにおいて、単語列に含まれることのある特定の記号に特化した第 3 の言語モデルをさらに参照することにより、前記単語列それぞれに付された出現確率の更新を行うことを特徴とする請求項 7 に記載の音声認識方法。

【請求項 11】 音声データを解析して特徴ベクトルに変換する音響処理ステップと、

前記音響処理ステップで得られた前記特徴ベクトルに対応する可能性がある音素列に、出現確率を付した音響データを生成する音響データ生成ステップと、

冗長語を単語候補に含めながら、前記音素列を単語列に変換する単語変換ステップと、

冗長語を含む単語列の出現する確率を連続するN個の単語の組み合わせと対応付けて記述する冗長語言語モデルを参照することにより、前記出現確率を更新する更新ステップと、

更新された前記出現確率が最も高い前記単語列を音声認識結果とする認識ステップと、

をコンピュータに実現させるためのプログラム。

【請求項12】 前記単語変換ステップは、冗長語として単語候補になった単語に他の単語との区別をするための記号を付して変換する請求項11に記載のプログラム。

【請求項13】 前記認識ステップは、出現確率が最も高い前記単語列をテキスト・データとして出力する請求項11に記載のプログラム。

【請求項14】 前記認識ステップは、出現確率が最も高い前記単語列から前記記号が付された前記単語を除外したテキスト・データとして出力する請求項12に記載のプログラム。

【請求項15】 前記更新ステップは、前記冗長語言語モデルの他に汎用的な言語モデルをも参照することにより前記出現確率を更新する請求項11に記載のプログラム。

【請求項16】 前記単語変換ステップにおいて、前記音声データに含まれるポーズ部分を句読点の候補として含めながら前記音素列を単語列に変換し、

前記更新ステップは、句読点挿入に限定した句読点言語モデルをさらに参照することにより前記出現確率を更新する、請求項11に記載のプログラム。

【請求項17】 音声データを解析して特徴ベクトルに変換する音響処理ステップと、

前記音響処理ステップで得られた前記特徴ベクトルに対応する可能性がある音素列に、出現確率を付した音響データを生成する音響データ生成ステップと、

冗長語を認識結果に反映させる場合には冗長語を単語候補に含めながら前記音素列を単語列に変換し、冗長語を認識結果に反映させない場合には冗長語を単語候補に含めずに前記音素列を単語列に変換する単語変換ステップと、

冗長語を認識結果に反映させる場合には冗長語を含む単語列の出現する確率を

連続するN個の単語の組み合わせと対応付けて記述する冗長言語モデルおよび汎用的な言語モデルを参照し、冗長語を認識結果に反映させない場合には汎用的な言語モデルを参照することにより、前記出現確率を更新する更新ステップと、

更新された前記出現確率が最も高い前記単語列を音声認識結果とする認識ステップと、

をコンピュータに実行させるためのプログラムを記録したコンピュータ読み取り可能な記録媒体。

【請求項 1 8】 前記単語変換ステップは、冗長語として単語候補になった単語に他の単語との区別をするための記号を付して変換する請求項 1 7 に記載の記録媒体。

【請求項 1 9】 前記冗長言語モデルおよび前記汎用的な言語モデルはN-gramモデルであり、

前記更新ステップは、前記冗長言語モデルおよび前記汎用的な言語モデルの加重平均値を用いて前記出現確率を更新する請求項 1 7 に記載の記録媒体。

【請求項 2 0】 音声認識結果に句読点を自動的に挿入させる場合に、前記単語変換ステップにおいて、前記音声データに含まれるポーズ部分を句読点の候補として含めながら前記音素列を単語列に変換し、

前記更新ステップは、句読点挿入に限定した句読点言語モデルをさらに参照することにより前記出現確率を更新する、請求項 1 7 に記載の記録媒体。

【発明の詳細な説明】

【0 0 0 1】

【発明の属する技術分野】

本発明は、人の自然な発話を認識して文章化し、冗長語（disfluency）と呼ばれる無意味な単語を自動的に除去してテキスト・データを作成することのできる音声認識装置およびその方法に関する。

【0 0 0 2】

【従来の技術】

例えば、「A Maximum Likelihood Approach to Continuous Speech Recognition (L.R. Bahl他, IEEE Trans. Vol. PAMI-5, No. 2, 1983, March)」や「単語

を認識単位とした日本語の大語彙連続音認識（西村他、情報処理学会論文誌、第40巻、第4号、1999年4月）」は、音響モデルと言語モデルを用いて音声認識を行う統計的方法を記述している。また、IBM ViaVoice98応用編（インフォ・クリエイツ出版事業部、1998年9月30日発行）の15ページには、言語モデルとして一般的な手法であるN-gram推定が開示されている。

さらに、応用上重要である自由発話の認識において頻出する「エー」「アノー」といった冗長語についても、これらをN-gramモデルの中で取り扱い、認識結果から自動的に除去する手法が「Statistical Language Modeling for Speech Disfluencies (A. Stolcke, E. Shriberg, Proc. of ICASSP96)」に開示されている。

【0003】

【発明が解決しようとする課題】

しかしながら、これらの手法を用いた場合、副作用として本来認識対象となるべき単語が冗長語であると判断されて、除かれてしまうという現象を避けることは難しい。冗長語の種類や頻度は、話者はもちろん、発話環境（例 原稿のあるなし、フォーマルかそうでないか）によっても変化するものであり、平均的なモデルでの予測を難しくしている。

したがって本発明は、ユーザの指定、または話者登録時に得られる情報から、冗長語の除去についてより適しているモデル設定機能を提供することを目的とする。

【0004】

【課題を解決するための手段】

語彙・言語モデルにおいてコンピュータなど特定分野に特化したもの（トピックと呼ばれる）を汎用の言語モデルと組み合わせて用い、当該分野における認識精度を向上させる手段が広く用いられている。トピックは比較的サイズを小さく作成することができ、そのオン／オフが容易である。そこで本発明では、冗長語除去のための語彙・言語モデルをトピックとして作成することに着目した。

つまり、本発明は、まず「アー」「エート」「ソノー」といった冗長語を通常

語同様に単語として登録した辞書と、その発音を追加した Base form Pool (ベースフォーム・プール) を用意する。ただし、冗長語については、後処理で除去する際に通常語と容易に区別できるよう、特別の記号 (たとえば不等号) を付加した表記 (<ア>、<エ>) としておくことが有効である。一方、冗長語を予測する専用の言語モデルをトピック用に作成する。この冗長語予測用の言語モデルは、冗長語を含むような N-gram および自由発話に特有の口語体表現に限って言語モデルの要素とするものであるから、汎用の言語モデルよりはるかに小さなものとすることができる。この言語モデルを汎用の言語モデルと次式 (1) で線形補間することにより、冗長語の予測精度を向上させる。

$$\text{Pr}(w_1|w_2, w_3) = \lambda P_1(w_1|w_2, w_3) + (1 - \lambda) P_2(w_1|w_2, w_3) \dots (1)$$

ただし、本式 (1) は $N = 3$ の場合を例示し、 $0 < \lambda \leq 1$ 、 P_1 は汎用言語モデルによる確率を、 P_2 は冗長語トピックの言語モデル (冗長語言語モデル) による確率を示す。

【0005】

本発明は以上の事項を基礎にするものであって、音声に含まれる音素列を 1 つ以上の単語列に変換し、変換して得られた前記単語列それぞれに対して、前記音素列により本来示される単語列であることの可能性を示す出現確率を付与する変換手段と、前記変換手段により得られた前記単語列それぞれに対応する言語モデルが示す更新用数値に基づいて、前記変換手段により前記単語列に付与された出現確率の更新を行う更新手段と、更新された出現確率が、前記音素列により本来示される単語列が最も高いことを示す前記単語列のいずれかを選択し、前記音声認識する認識手段とを備え、前記更新手段は、自由発話特有の表現に特化した第 1 の言語モデルと、前記第 1 の言語モデルとは異なる第 2 の言語モデルとを用いて前記更新用数値を算出し、算出した前記更新用数値に基づいて、前記出現確率の更新を行うことを特徴とする音声認識装置である。

冗長語は、自由発話特有の表現に関する単語組に含まれる。したがって、本発明は、冗長語に限定することなく、自由発話特有の表現を対象とする第 1 の言語

モデルを備える。自由発話特有の表現としては、例えば、「なん」「です」「ね」がある。

【0006】

本発明の音声認識装置において、前記第1の言語モデルは、自由発話特有の表現に含まれる所定の単語を含む単語列が、前記音素列により本来示される単語列である確からしさを示すものとすることができる。

また本発明の音声認識装置において、前記所定の単語を音声認識の結果に含める場合には、前記変換手段は前記音素列を前記所定の単語を含む単語列に変換し、前記更新手段は、前記第1の言語モデルおよび前記第2の言語モデルに基づいて、前記単語列それぞれの出現確率を更新することができる。

さらに本発明において、前記第1の言語モデルは、冗長語を含む単語組みを要素とすることができる。

また本発明の音声認識装置において、前記第1の言語モデルおよび前記第2の言語モデルをN-gramモデルとし、前記更新手段は、前記第1の言語モデルおよび前記第2の言語モデルの加重平均値を前記更新用数値として用いることができる。

本発明は、コンピュータ・システム内に以上の音声認識装置を備えることができる。

【0007】

本発明はまた、以下の音声認識方法を提供する。すなわち本発明の音声認識方法は、音声に含まれる音素列を1つ以上の単語列に変換し、変換して得られた前記単語列それぞれに対して、前記音素列により本来示される単語列であることの可能性を示す出現確率を付与する変換ステップと、前記変換ステップにより得られた前記単語列が自由発話固有の単語を含む場合に、自由発話固有の表現に特化した第1の言語モデルおよび前記第1の言語モデルとは異なる第2の言語モデルを参照することにより、前記単語列それぞれに付された出現確率の更新を行う更新ステップと、更新された出現確率が、前記音素列により本来示される単語列が最も高いことを示す前記単語列のいずれかを選択し、前記音声を認識する認識ステップと、を備えたことを特徴とする。

【 0 0 0 8 】

本発明の音声認識方法において、前記第 1 の言語モデルは、自由発話特有の単語を含む単語列の出現確率を、連続する N 個の単語の組み合わせと対応付けて記述することができる。

また本発明の音声認識方法において、前記自由発話固有の単語の典型例として冗長語がある。

さらに本発明の音声認識方法では、前記更新ステップにおいて、単語列に含まれることのある特定の記号に特化した第 3 の言語モデルをさらに参照することにより、前記単語列それぞれに付された出現確率の更新を行うこともできる。「ある特定の記号」は、句点、読点のほか？といった記号を包含する。したがって、これらの特定の記号を自動的に挿入することを可能とする。

【 0 0 0 9 】

本発明の音声認識方法は、コンピュータで所定の処理を実行するためのプログラムとしても成立する。すなわち本発明は、音声データを解析して特徴ベクトルに変換する音響処理ステップと、前記音響処理ステップで得られた前記特徴ベクトルに対応する可能性がある音素列に、出現確率を付した音響データを生成する音響データ生成ステップと、冗長語を単語候補に含めながら、前記音素列を単語列に変換する単語変換ステップと、冗長語を含む単語列の出現する確率を連続する N 個の単語の組み合わせと対応付けて記述する冗長語言語モデルを参照することにより、前記出現確率を更新する更新ステップと、更新された前記出現確率が最も高い前記単語列を音声認識結果とする認識ステップと、をコンピュータに実現させるためのプログラムである。

【 0 0 1 0 】

本発明のプログラムにおいて、前記単語変換ステップは、冗長語として単語候補になった単語に他の単語との区別をするための記号を付して変換することができる。冗長語であることを明示させるため、あるいは冗長語を自動削除する際の日印とするためである。

また本発明のプログラムにおいて、前記認識ステップは、出現確率が最も高い前記単語列をテキスト・データとして出力する。テキスト・データとして出力す

る際に、出現確率が最も高い前記単語列から前記記号が付された前記単語を除外したテキスト・データとすることもできる。冗長語を自動的に削除してテキスト・データを表示する際に有効である。

本発明のプログラムにおいて、前記更新ステップは、前記冗長語言語モデルの他に汎用的な言語モデルをも参照することにより前記出現確率を更新することが実用上は必要になってくる。

さらに本発明のプログラムでは、前記単語変換ステップにおいて、前記音声データに含まれるポーズ部分を句読点の候補として含めながら前記音素列を単語列に変換し、前記更新ステップは、句読点挿入に限定した句読点言語モデルをさらに参照することにより前記出現確率を更新することもできる。句読点を自動的に挿入する際に有効となる。

【0011】

本発明の音声認識方法は、コンピュータに実行させるためのプログラムを記録したコンピュータ読み取り可能な記録媒体としても成立する。すなわち本発明の記録媒体は、音声データを解析して特徴ベクトルに変換する音響処理ステップと、前記音響処理ステップで得られた前記特徴ベクトルに対応する可能性がある音素列に、出現確率を付した音響データを生成する音響データ生成ステップと、冗長語を認識結果に反映させる場合には冗長語を単語候補に含めながら前記音素列を単語列に変換し、冗長語を認識結果に反映させない場合には冗長語を単語候補に含めずに前記音素列を単語列に変換する単語変換ステップと、冗長語を認識結果に反映させる場合には冗長語を含む単語列の出現する確率を連続するN個の単語の組み合わせと対応付けて記述する冗長語言語モデルおよび汎用的な言語モデルを参照し、冗長語を認識結果に反映させない場合には汎用的な言語モデルを参照することにより、前記出現確率を更新する更新ステップと、更新された前記出現確率が最も高い前記単語列を音声認識結果とする認識ステップと、をコンピュータに実行させる。

【0012】

本発明の記録媒体において、前述した本発明のプログラムと同様に、単語変換ステップは、冗長語として単語候補になった単語に他の単語との区別をするため

の記号を付して変換することができる。

また、本発明の記録媒体において、前記冗長言語モデルおよび前記汎用的な言語モデルはN-gramモデルであり、前記更新ステップは、前記冗長言語モデルおよび前記汎用的な言語モデルの加重平均値を用いて前記出現確率を更新することができる。

さらに、本発明の記録媒体において、音声認識結果に句読点を自動的に挿入させる場合に、前記単語変換ステップにおいて、前記音声データに含まれるポーズ部分を句読点の候補として含めながら前記音素列を単語列に変換し、前記更新ステップは、句読点挿入に限定した句読点言語モデルをさらに参照することにより前記出現確率を更新することが有効である。

【0013】

【発明の実施の形態】

以下本発明を実施の形態に基づき説明する。

＜第1の実施形態＞

図1は第1の実施形態による音声認識装置1の構成を示す。

図1に示すように、音声認識装置は1、マイクロプロセッサ、メモリおよびこれらの周辺回路などから構成されるCPU10、マウス、キーボード、およびマイクロフォンなどの音声入力装置120を含む入力装置12、CRTディスプレイなどの表示装置14、HDD (Hard Disk Drive)、DVD (Digital Versatile Disc) 装置、CD (Compact Disk) 装置などの記憶装置16から構成される。

つまり、音声認識装置1は、通常のコンピュータに音声入力装置120を付加したハードウェア構成を採り、DVD、CD-ROMあるいはCD-R等の記憶媒体18に記憶された形態で供給される音声認識プログラム2aを実行する。つまり、音声認識装置1は、音声入力装置120から入力され、デジタル形式に変換された音声（音声データ）を認識して、冗長語を自動的に除去したテキスト・データを生成し、記憶装置16に挿入された記憶媒体18に記憶し、あるいは表示装置14にテキスト・データを表示する。

【0014】

図 2 は音声認識プログラム 2 a の構成を示す図である。

図 2 に示すように、音声認識プログラム 2 a は、音響処理部 2 0、言語復号部 2 2、音響モデル 3 0、言語モデル 3 2 a、ベースフォーム・プール 3 4 a および言語モデル 3 2 a についての ON/OFF 制御部 3 6 から構成される。言語モデル 3 2 a は、汎用言語モデル 3 2 0 と冗長語言語モデル（トピック） 3 2 2 とを含んでいる。

【 0 0 1 5 】

音響処理部 2 0 は、音声入力装置 1 2 0 から入力された音声データに対してフーリエ変換などの周波数分析処理および特徴抽出処理を行い、特徴ベクトルに変換し、言語復号部 2 2 に対して出力する。

音響モデル 3 0 は、音素毎に特徴ベクトルの出現確率（音素列により本来示される単語列である可能性が高ければ高いほど高い数値を示す）を HMM（Hidden Markov Model）等により記述した音響モデルデータを言語復号部 2 2 に対して出力する。

【 0 0 1 6 】

ベースフォーム・プール 3 4 a は、言語モデル 3 2 a の汎用言語モデル 3 2 0 に対応する汎用部分と、冗長語言語モデル 3 2 2 に対応する冗長語発音（例「エー」「アー」）部分とを有する。また、ベースフォーム・プール 3 4 a は、各単語の発音を音響モデル 3 0 に対応付けられた発音記号で記述する。

【 0 0 1 7 】

言語モデル 3 2 a において、汎用言語モデル 3 2 0 は、音声認識に必要な一般的な単語・単語列の出現確率を、連続する N 個（例えば、N = 3）の単語の組み合わせとし、出現確率と対応付けて表形式で記述する。なお、この汎用言語モデル 3 2 0 のサイズは、通常、6 0 M B 程度になる。ここで、N = 3 個の場合の単語組み合わせの出現確率は、連続する 2 個の単語列の後にある 1 個の単語が出現する確率として定義され、汎用言語モデル 3 2 0 においては、冗長語は認識対象単語とみなされない。

【 0 0 1 8 】

冗長語言語モデル 3 2 2 には、本モデル学習のため自由発話を書き起こしたテ

キストを走査することにより作成され、冗長語を含む単語組（N = 3 の例では「これ」「が」「<エー>」）や自由発話特有の口語的表現に関する単語組（例「なん」「です」「ね」）が含まれる。冗長語も自由発話特有の表現の1つである。後者は冗長語検出と直接関連するわけではないが、汎用言語モデル 3 2 0 の多くが新聞などの「書き言葉」から学習されたものであるため、自由発話の認識精度を全体としてより向上させる効果をもつ。

【 0 0 1 9 】

図 3 は、図 2 に示した音声認識プログラム 2 a（言語復号部 2.2）の動作を示す図である。図 3（a）は冗長語言語モデル 3 2 2 が用いずに汎用言語モデル 3 2 0 のみを用いて音声認識が行われる（トピック・オフの）場合を示し、（b）は汎用言語モデル 3 2 0 および冗長語言語モデル 3 2 2 を用いて音声認識が行われる（トピック・オンの）場合を示す。

言語復号部 2.2 は、音響処理部 2.0 から入力された特徴ベクトルを、音響モデル 3.0、ベースフォーム・プール 3.4 a および言語モデル 3.2 a により計算される確率が最大になる単語列（下記式（2）の W' ）をテキスト・データとして、表示装置 1.4 または記憶装置 1.6 に対して出力し、これらに表示あるいは記憶する。

【 0 0 2 0 】

言語復号部 2.2 は、図 3（a）、（b）に例示するように、冗長語言語モデル 3 2 2 が用いられるか否かに応じて、図 4～図 6 を参照して後述するように、冗長語の検出と削除を自動的に行う。

ON/OFF 制御部 3.6 は冗長語言語モデル 3.2.2 の ON/OFF を制御し、冗長語言語モデル 3.2.2 の使用を決定する。この制御方式にはさまざまなものが考えられるが、もっとも簡単なものはユーザからの指示に応じて冗長語言語モデル 3.2.2 の ON/OFF を決定するものである。一方、より自動的に制御するには、ユーザの音声データの一部を冗長語言語モデル 3.2.2 が ON の場合と、OFF の場合について 2 度言語復号を行い、その結果得られるスコア（ゆう度）を比較することによって、いずれがより適しているかを判断し、ON/OFF を制御すればよい。この判断は使用するたびに行うことも可能であるが、言語復号を 2

度行うことはそれだけCPU10の演算時間を必要とするので、通常エンロールと呼ばれるユーザ登録時に行うことが望ましい。

また、下式(2)において、 $P(S|W)$ は、音響モデル30およびベースフォーム・プール34aにより与えられ、 $P(W)$ は言語モデル32aにより与えられる。この $P(W)$ は、上記式(1)により、汎用言語モデル320と冗長言語モデル322とが λ の値に応じて加重平均され、求められる。 λ の値を変更することで、冗長言語モデル322を単にON/OFFするのみならず、汎用言語モデル320および冗長言語モデル322のいずれのモデルをどの程度重視するかの制御が可能となる。一般にはシステム設計者が、事前の実験により、最適と考えられる値に設定するが、これについてもエンロール時の試験的な言語復号により、調整することが可能である。

【0021】

$$W' = \operatorname{argmax}_W P(W|S) = \operatorname{argmax}_W P(S|W)P(W) \dots (2)$$

ただし、 S は特徴ベクトル(s_1, s_2, \dots, s_k)、 W は単語列(w_1, w_2, \dots, w_l)、 W' は単語列(w'_1, w'_2, \dots, w'_l)である。

【0022】

以下、図4～図6をさらに参照し、N-gramとして3-gramを用いる場合を具体例として、言語復号部22の動作を詳細に説明する。

図4は、図2に示した音声認識プログラム2a(言語復号部22)の音声認識処理を示すフローチャートである。

図5は、図2に示した汎用言語モデル320のみを用いるため、冗長語の自動削除を行わない(トピック・オフ)場合の音声認識プログラム2a(言語復号部22)の動作を示す図である。

図6は、図2に示した汎用言語モデル320および冗長言語モデル322を用いて冗長語の自動削除を行う(トピック・オン)場合の音声認識プログラム2a(言語復号部22)の動作を示す図である。

【0023】

図4に示すように、ステップ100（図中S100、以下同様）において、音響処理部20が、入力音声としての「これがえほん」を特徴ベクトルに変換し、言語復号部22に対して出力する。

ステップ102（S102）において、図5および図6に示すように、言語復号部22は、入力された特徴ベクトル（ s_1, s_2, \dots, s_k ）を音響モデル30に対して出力する。

音響モデル30は、特徴ベクトルに対応する可能性がある音素列（/koegaehon/, /koregaehon/）に出現確率を付した音響データを作成し、言語復号部22に返す。

【0024】

ステップ104（S104）において、言語復号部22は、ユーザが指定するか、またはエンロール時の試験復号により冗長語自動削除機能が設定されているか否か、つまり冗長語言語モデル322（トピック）がオンになっているか否かを判断し、自動削除が行われない設定になっている（トピックがオフになっている）場合には λ を1に設定してステップ110（S110）の処理に進み、そうでない場合は0～1のあらかじめ決められた値に設定し、ステップ120（S120）の処理に進む。

【0025】

冗長語の自動削除を行わない場合、ステップ110（S110）において、図5に示すように、言語復号部22は、ベースフォーム・プール34aの汎用部分（汎用ベースフォーム・プール）のみを参照して、冗長語を候補に含めずに、順次音素列（/koegaehon/, /koregaehon/等）を単語列（声/が/絵本、これ/が/絵本等）に変換し、候補とする。すなわち/ehon/の部分はベースフォーム/ehon/に該当する通常単語、もしくは通常単語の組み合わせに変換されることになる。ただし、この段階での変換は音響モデル30から得られる確率のみを用いて行われる。

ステップ112（S112）において、図5に示すように、言語復号部22は、汎用言語モデル320のみを参照して、ステップ110（S110）の処理において得た候補単語列の出現確率を更新する。図5の例においては、この処理の

結果として、単語列「これ/が/絵本」の確率が0.02、単語列「声/が/絵本」の確率が0.01となる。

【0026】

冗長語の自動削除を行う場合、ステップ120 (S120) において、図6に示すように、言語復号部22は、ベースフォーム・プール34aの汎用ベースフォーム・プールおよび冗長語の音に対応するよう記述された冗長語部分（冗長語ベースフォーム・プール）の両方を参照して、順次冗長語を単語候補に含めながら、音素列（/koegaehon/, /koregaehon/等）を単語列（声/が/絵本、声/が/<エ>/本、これ/が/<エ>/本 等）に変換する。なお、<エ>というように、エに<>を付しているのは、それが冗長語であることを明示するためである。また、この<>を付した単語を自動的に削除したテキスト・データを表示させる場合に有効である。

【0027】

ステップ122 (S122) において、図6に示すように、言語復号部22は、汎用言語モデル320および冗長語言語モデル322を、 $\lambda \neq 1$ とした前記式(1)にしたがって重み付けして参照し、S110の処理において得た候補単語列の出現確率を更新する。図6の例においては、この処理の結果として、単語列「これ/が/<エ>/本」の確率が0.02、その他の単語列の確率が0.01となる。

【0028】

ステップ124 (S124) において、図5, 6に示すように、言語復号部22は、S112またはS122の処理において計算・更新された確率が最も大きい単語列を、音声認識結果を示すテキスト・データとして出力する。たとえば図6の例では「これ/が/<エ>/本」が選択されることになる。なお、図6では<エ>を表示しているが、実際には<エ>を削除した状態でテキスト・データを表示させることもできる。以下の例でも同様であり、見かけ上<エ>を表示しているが、実際には表示されていない場合を包含する意味である。

【0029】

冗長語言語モデル322を用いない場合と用いる場合に分けて、話者が、例え

ば、音声認識装置 1 の音声入力装置 120 (図 1) から、「コレガ」「エ」「ホン」という音声を入力した場合の動作を説明する。

冗長語言語モデル 322 を用いない場合、音響処理部 20 は、この音声を示す音声データを処理し、この音声の音を記述する特徴ベクトルを言語復号部 22 に対して出力する。言語復号部 22 は、図 4 および図 5 に示したように、音響モデル 30 と汎用言語モデル 320 のみを用いて音響処理部 20 から入力された特徴ベクトルを処理し、音響モデル 30 から得られた確率と同言語モデル 32a から得られた確率を評価することにより「コレガ」を「これ/が」と識別し、「エ」

「ホン」を「絵本」と識別する。すなわち「エ」という音に対しては、通常語の組み合わせの中から、音響モデル 30 および言語モデル 32a を参照して得られる確率が最大になるものを選択し、この場合「これ/が/絵本」というテキスト・データを識別結果として出力する。

【0030】

冗長語言語モデル 322 を用いる場合、冗長語言語モデル 322 を用いない場合と同じく、当該音声データを処理し、この音声の音を記述する特徴ベクトルを言語復号部 22 に対して出力する。言語復号部 22 は、図 4 および図 6 に示したように、音響モデル 30、汎用言語モデル 320、および冗長語言語モデル 322 を用いて音響処理部 20 から入力された特徴ベクトルを処理し、音響モデル 30 から得られた確率と同言語モデル 32a から得られた確率を評価することにより「コレガ」を「これ/が」と識別し、「エ」「ホン」を「<エ>」「本」と識別する。すなわち「エ」という音に対しては通常語である「絵」などの他に冗長語「<エ>」の可能性も考慮した上で言語モデル 32a から得られる確率により、文脈上いずれがより高い可能性を持つかが判定される。そして図 3 に示すような状況においては、「これ/が/絵本/発明」と「これ/が/<エ>/本/発明」という 2 つの単語列を比較することで、後者がより高い確率であることがより一層容易に理解される。

【0031】

冗長語言語モデル 322 は、冗長語にそれを示す記号を割り当てた発音辞書などから構成され、上述した冗長語自動削除用のタスクに比べて、非常に小さいサ

イズ（例えば、2MB程度）で済む。

また、本実施の形態による冗長語自動削除方法では、トピックとしての冗長語言語モデル322を追加すれば足りるので、従来からの汎用のシステムをほとんど変更することなく、冗長語自動削除機能を追加することができる。

しかもユーザは、この冗長語自動削除用の冗長語言語モデル322を選択するか否かによって、容易に冗長語自動削除機能をオン・オフすることができ、かつ、このオン・オフにはプログラムの再起動が不要である。

さらに、冗長語言語モデル322を、例えば、「コンピュータ」、「サッカー」、「野球」といった、他の分野に特化されたトピックと組み合わせて使用することも可能である。例えば、句読点等の記号を挿入するためのトピックを併設することが有効である。以下、冗長語削除用トピックと句読点挿入用トピックとを併設した例を第2の実施形態として説明する。

【0032】

<第2の実施形態>

第2の実施形態による音声認識装置の基本的な構成は、図1に示した第1の実施形態による音声認識装置1と一致する。したがって、ここでの説明は省略する。

図7は第2の実施形態による音声認識プログラム2bの構成を示す図である。音声認識プログラム2bの基本的な機能は、第1の実施形態による音声認識プログラム2aと一致している。したがって、一致する部分については、図7に第1の実施形態による音声認識プログラム2aと同一の符号を付している。

音声認識プログラム2bは、言語モデル32bが、汎用言語モデル320および冗長語言語モデル322の他に、さらに句読点言語モデル323を備えている点で相違する。また、ベースフォーム・プール34bは、句読点言語モデル323に対応し、音響モデルデータ中の空白部分（ポーズ部分；句読点「。」、「」に対応する部分）を検出する句読点部分とを有する点でも相違する。なお、図7において省略しているベースフォーム・プール34bの内容は後述する。

【0033】

また、言語復号部22において、句読点自動挿入を行う場合には音声のポーズ

部分を単語とみなして処理を行い、句読点自動挿入を行わない場合には音声のポーズ部分を単語としてみなさずに処理を行う。

【 0 0 3 4 】

言語モデル 3 2 b において、汎用言語モデル 3 2 0 および冗長言語モデル 3 2 2 の機能は第 1 の実施形態と同様であるので、ここでは、句読点言語モデル 3 2 3 についてのみ説明する。

句読点言語モデル 3 2 3 は、句読点挿入のために特化したトピックであって、句読点挿入に必要な特定の単語・単語列の出現確率を、連続する 3 個の単語の組み合わせとし、出現確率と対応付けて表形式に記述する。

句読点言語モデル 3 2 3 は、汎用言語モデル 3 2 0 を、句読点を単語とみなすようにして、ベースフォーム・プール 3 4 b によりポーズと判断された部分に句読点を自動的に挿入するために特化させ、データ量を少なくしたものであると考えることができる。

【 0 0 3 5 】

句読点言語モデル 3 2 3 には、句読点クラスとの相互情報量に基づいて選択された単語、例えば、以下に例示するような句点・読点の直前に位置する単語が含まれる。なお、以下の例示は、上位 2 0 語を列挙するものであり、また、H m は句読点クラスとの相互情報量の値を示す。

なお、句読点言語モデル 3 2 3 は、内容を句読点挿入に必要な情報に限定するので、一般に、データ量を汎用言語モデル 3 2 0 の $1/100 \sim 1/1000$ 程度とすることができる。

【 0 0 3 6 】

句読点言語モデル 3 2 3 に含まれる単語

H m	単語
2 7 5 . 1 1 1	いる
1 9 7 . 1 6 6	だ
1 6 0 . 2 2 3	した
1 5 9 . 4 2 5	です
1 5 2 . 8 8 9	は

137.400	し
137.164	ね
129.855	で
112.604	ある
103.377	が
79.751	ます
73.160	か
66.952	しかし
65.562	おり
63.930	ので
63.078	ました
62.469	います
59.100	だが
49.474	ない
48.714	では

【0037】

図8は、図7に示した音声認識プログラム2b（言語復号部22）の動作を示す図である。図8（a）は汎用言語モデル320のみを用いて音声認識が行われる場合を示し、（b）は汎用言語モデル320および句読点言語モデル323を用いて音声認識が行われる場合を、（c）は汎用言語モデル320および冗長言語モデル322を用いて音声認識が行われる場合を、さらに（d）は汎用言語モデル320、冗長言語モデル322および句読点言語モデル323を用いて音声認識が行われる場合を示す。

【0038】

言語復号部22は、音響処理部20から入力された特徴ベクトルを、音響モデル30、ベースフォーム・プール34bおよび言語モデル32bにより計算される確率が最大になる単語列（上記式（2）の W' ）をテキスト・データとして、表示装置14または記憶装置16に対して出力し、これらに表示あるいは記憶する。

また、言語復号部 2 2 は、図 8 (a) ~ (d) に例示するように、冗長語言語モデル 3 2 2、句読点言語モデル 3 2 3 が用いられるか否かに応じて、図 9 ~ 図 1 3 を参照して後述するように、冗長語の検出と削除、句読点の挿入を自動的に行う。

上記式 (2) において、 $P(S|W)$ は、音響モデル 3 0 およびベースフォーム・ルール 3 4 b によって与えられ、 $P(W)$ は言語モデル 3 2 b によって与えられる。この $P(W)$ は、上記式 (1) により、汎用言語モデル 3 2 0、冗長語言語モデル 3 2 2 および句読点言語モデル 3 2 3 とが λ の値に応じて加重平均され、求められる。 λ の値を変更することで、冗長語言語モデル 3 2 2、句読点言語モデル 3 2 3 を単に ON/OFF するのみならず、汎用言語モデル 3 2 0、冗長語言語モデル 3 2 2 および句読点言語モデル 3 2 3 のいずれのモデルをどの程度重視するかの制御が可能となる。一般にはシステム設計者が、事前の実験により、最適と考えられる値に設定するが、これについてもエンロール時の試験的な言語復号により、調整することが可能である。

【 0 0 3 9 】

以下、図 9 ~ 図 1 3 を参照して N - g r a m として 3 - g r a m を用いる場合を具体例として、言語復号部 2 2 の動作を詳細に説明する。

図 9 は、図 7 に示した音声認識プログラム 2 b (言語復号部 2 2) の処理を示すフローチャートである。

図 1 0 は、図 7 に示した汎用言語モデル 3 2 0 のみを用いるため、冗長語の自動削除および句読点の自動挿入を行わない (2 つのトピック・オフ) の場合の音声認識プログラム 2 b (言語復号部 2 2) の動作を示す図である。

図 1 1 は、図 7 に示した汎用言語モデル 3 2 0 と句読点言語モデル 3 2 3 を用いて句読点の自動挿入を行う (句読点トピック・オン) 場合の音声認識プログラム 2 b (言語復号部 2 2) の動作を示す図である。

図 1 2 は、図 7 に示した汎用言語モデル 3 2 0 と冗長語言語モデル 3 2 2 を用いて冗長語の自動削除を行う (冗長語トピック・オン) 場合の音声認識プログラム 2 b (言語復号部 2 2) の動作を示す図である。

図 1 3 は、図 7 に示した汎用言語モデル 3 2 0、冗長語言語モデル 3 2 2 およ

び句読点言語モデル 3 2 3 を用いて、冗長語の自動削除（冗長語トピック・オン）および句読点の自動挿入（句読点トピック・オン）を行う場合の音声認識プログラム 2 b（言語復号部 2 2）の動作を示す図である。

【 0 0 4 0 】

図 9 に示すように、ステップ 2 0 0（図中 S 2 0 0、以下同様）において、音響処理部 2 0 が、入力音声としての「これがえほん」を特徴ベクトルに変換し、言語復号部 2 2 に対して出力する。

ステップ 2 0 2（S 2 0 2）において、図 1 0～図 1 3 に示すように、言語復号部 2 2 は、入力された特徴ベクトル（ s_1, s_2, \dots, s_k ）を音響モデル 3 0 に対して出力する。

音響モデル 3 0 は、特徴ベクトルに対応する可能性がある音素列（/koegaehon /、/koregaehon/）に出現確率を付した音響データを作成し、言語復号部 2 2 に返す。

【 0 0 4 1 】

ステップ 2 0 4（S 2 0 4）において、言語復号部 2 2 は、ユーザが指定するか、またはエンロール時の試験復号により冗長語自動削除機能が設定されているか否か、つまり冗長語言語モデル 3 2 2（冗長語トピック）がオンになっているか否かを判断し、自動削除が行われない設定になっている（冗長語トピックがオフになっている）場合には、 λ を 1 に設定してステップ 2 0 8（S 2 0 8）の判断に進み、そうでない場合は、0～1 λ のあらかじめ決められた値に設定し、ステップ 2 0 6（S 2 0 6）の判断に進む。

ステップ 2 0 6（S 2 0 6）において、言語復号部 2 2 は、ユーザが指定するか、またはエンロール時の試験復号により句読点自動挿入機能が設定されているか否か、つまり句読点言語モデル 3 2 3（句読点トピック）がオンになっているか否かを判断し、自動挿入が行われない設定になっている（句読点トピックがオフになっている）場合には、 λ を 1 に設定してステップ 2 2 0（S 2 2 0）の処理に進み、そうでない場合は、0～1 λ のあらかじめ決められた値に設定し、ステップ 2 1 0（S 2 1 0）の処理に進む。ステップ 2 1 0 は冗長語トピックおよび句読点トピックの両者がオンの場合の処理であり、ステップ 2 2 0 は冗長語ト

ピックがオンの場合の処理である。

【0042】

ステップ208 (S208) において、言語復号部22は、ステップ206 (S206) と同様に、句読点言語モデル323 (句読点トピック) がオンになっているか否かを判断し、自動挿入が行われない設定になっている (句読点トピックがオフになっている) 場合には、λを1に設定してステップ240 (S240) の処理に進み、そうでない場合は、0～1λのあらかじめ決められた値に設定し、ステップ230 (S230) の処理に進む。ステップ230は句読点トピックがオンの場合の処理であり、ステップ240は冗長語トピックおよび句読点トピックの両者がオフの場合の処理である。

【0043】

ステップ240 (S240) において、図10に示すように、言語復号部22は、ベースフォーム・プール34bの汎用部分 (汎用ベースフォーム・プール) のみを参照して、冗長語およびポーズを単語候補に含めずに、順次、音素列 (/koegatenehon/, /koregatenehon/ 等) を単語列 (声/が/, /絵本, これ/が/, /絵本 等) に変換し、候補とする。すなわち /ehon/ の部分はベースフォーム /ehon/ に該当する通常単語、もしくは通常単語の組み合わせに変換されることになる。ただし、この段階での変換は音響モデル30から得られる確率によってのみ行われる。

ステップ242 (S242) において、図10に示すように、言語復号部22は、汎用言語モデル320のみを参照して、ステップ220 (S220) の処理において得た候補単語列の出現確率を更新する。図10の例においては、この処理の結果として、単語列「これ/が/, /絵本」の確率が0.02、単語列「声/が/, /絵本」の確率が0.01となる。

【0044】

ステップ230 (S230) において、図11に示すように、言語復号部22は、ベースフォーム・プール34bの汎用ベースフォーム・プールおよびポーズを検出する句読点部分 (句読点トピック) の両方を参照して、順次、ポーズを単語に含めて、音素列 (/koegaehon/, /koregaehon/ 等) を単語列 (/声/が/絵本/,

/これ/が/え/本/, /声/が/, /絵本/, /これ/が/, /え/本/ 等) に変換する。

ステップ232 (S232) において、図11に示すように、言語復号部22は、汎用言語モデル320および句読点言語モデル323を、 $\lambda \neq 1$ とした前記式(1)にしたがって重み付けして参照し、S230の処理において得た候補単語列の出現確率を更新する。図11の例においては、この処理の結果として、単語列「/これ/が/, /え/本/」の確率が0.02、その他の単語列の確率が0.01となる。

【0045】

ステップ220 (S220) において、図12に示すように、言語復号部22は、ベースフォーム・プール34bの汎用ベースフォーム・プールおよび冗長語の音に対応するよう記述された冗長語部分(冗長語トピック)の両方を参照して、順次、冗長語を単語候補に含めながら、音素列(/koegatenehon/, /koregatenehon/等)を単語列(声/が/, /絵本, これ/が/, /絵本, これ/が/, /<エ>/本 等)に変換する。トピック・オフの場合と同様この段階での変換は音響モデル30から得られる確率によってのみ行われる。

ステップ222 (S222) において、図12に示すように、言語復号部22は、汎用言語モデル320および冗長語言語モデル322を、 $\lambda \neq 1$ とした前記式(1)にしたがって重み付けして参照し、S220の処理において得た候補単語列の出現確率を更新する。図12の例においては、この処理の結果として、単語列「これ/が/, /<エ>/本」の確率が0.02、その他の単語列の確率が0.01となる。

【0046】

ステップ210 (S210) において、図13に示すように、言語復号部22は、ベースフォーム・プール34bの汎用ベースフォーム・プール、冗長語部分(冗長語トピック)、および句読点部分(句読点トピック)の両方を参照して、順次、冗長語およびポーズを単語候補に含めながら、音素列(/koegaehon/, /koregaehon/等)を単語列(声/が/絵本, 声/が/<エ>/本, これ/が/, /<エ>/本 等)に変換する。

ステップ212 (S212) において、図13に示すように、言語復号部22

は、汎用言語モデル 3 2 0、冗長言語モデル 3 2 2 および句読点言語モデル 3 2 3 を、 $\lambda \neq 1$ とした前記式 (1) にしたがって重み付けして参照し、S 2 1 0 の処理において得た候補単語列の出現確率を更新する。図 1 3 の例においては、この処理の結果として、単語列「これ/が/、/<エ>/本」の確率が 0. 0 2、その他の単語列の確率が 0. 0 1 となる。

【 0 0 4 7 】

ステップ 2 2 4 (S 2 2 4) において、図 1 0 ~ 図 1 3 に示すように、言語復号部 2 2 は、順次、S 2 1 2, S 2 2 2, S 2 3 2 および S 2 4 2 の処理において更新された出現確率が最も高い単語列を、音声認識結果を示すテキスト・データとして出力する。

【 0 0 4 8 】

以下、第 2 の実施形態による音声認識装置 1 (図 1, 図 7) の動作を図 8 に基づき説明する。

図 8 (a) に示すように、冗長言語モデル 3 2 2 および句読点言語モデル 3 2 3 を用いない場合 ($\lambda = 1$) には、話者が、例えば、音声認識装置 1 の音声入力装置 1 2 0 (図 1) から、「コレガ」・「p a u s e (無音を示す。以下、同じ)」・「テン」・「エ」・「ホンハツメイノ」・「p a u s e」・「ヨウテン」という音声を入力すると、音響処理部 2 0 は、この音声を示す特徴ベクトルを言語復号部 2 2 に対して出力する。

【 0 0 4 9 】

言語復号部 2 2 (図 2) は、図 7 および図 1 0 に示したように、音響モデル 3 0 と汎用言語モデル 3 2 0 のみを用いて音響処理部 2 0 から入力された特徴ベクトルを処理し、音響モデル 3 0 から得られた確率と汎用言語モデル 3 2 0 から得られた確率を評価することにより「コレガ」を「これ/が」と識別し、「p a u s e」に続く「テン」を読点「、」と識別する。つづいて、「エ」「ホン」を「絵本」と識別し、「ハツメイノ」を「発明の」と識別する。さらに、言語復号部 2 2 は、「ホンハツメイノ」のあとの「p a u s e」には「テン」、「マル」が続いていないので、その後の「ヨウテン」を「要点」と識別して、「これが、絵本発明の要点」というテキスト・データを識別結果として出力する。

【0050】

図8(b)に示すように、冗長言語モデル322を用いない場合($\lambda = 1$)には、話者が、例えば、音声認識装置1の音声入力装置120(図1)から、「コレガ」・「pause」・「エ」・「ホンハツメイノ」・「pause」・「ヨウテン」という音声を入力すると、音響処理部20は、この音声を示す音素で示す特徴ベクトルを言語復号部22に対して出力する。

【0051】

言語復号部22(図2)は、図7および図11に示したように、音響モデル30、汎用言語モデル320、および句読点言語モデル323を用いて音響処理部20から入力された特徴ベクトルを処理し、音響モデル30から得られた確率と汎用言語モデル320および句読点言語モデル323から得られた確率を評価することにより「コレガ」を「これ/が」と識別し、「ガ」に続く「pause」を読点「、」と識別する。つづいて、「エ」「ホン」を「絵本」と識別し、「ハツメイノ」を「発明の」と識別する。さらに、「ハツメイノ」には、「pause」が続くが、「ノ」の後ろには通常、句読点は続かないので、言語復号部22は、この「pause」の部分には句読点を挿入せず、その後の「ヨウテン」を「要点」と識別して、「これが、絵本発明の要点」というテキスト・データを識別結果として出力する。

【0052】

図8(c)に示したように、句読点言語モデル323を用いない場合($\lambda = 1$)には、話者が、例えば、音声認識装置1の音声入力装置120(図1)から、「コレガ」・「pause」・「テン」・「エ」・「ホンハツメイノ」・「pause」・「ヨウテン」という音声を入力すると、音響処理部20は、この音声を示す特徴ベクトルを言語復号部22に対して出力する。

【0053】

言語復号部22(図2)は、図7および図12に示したように、汎用言語モデル320および冗長言語モデル322を用いて音響処理部20から入力された特徴ベクトルを処理し、音響モデル30から得られた確率と汎用言語モデル320および冗長言語モデル322から得られた確率を評価することにより「コレ

ガ」を「これ/が」と識別し、「p a u s e」に続く「テン」を読点「、」と識別する。つづいて、「エ」「ホン」を「<エ>」「本」と識別する。すなわち「エ」という音に対しては通常語である「絵」などの他に冗長語「<エ>」の可能性も考慮した上で言語モデル32bから得られる確率により、文脈上いずれがより高い可能性を持つかが判定される。さらに、言語復号部22は、「ホンハツメイノ」のあとの「p a u s e」には「テン」、「マル」が続いていないので、その後の「ヨウテン」を「要点」と識別して、「これが、<エ>本発明の要点」というテキスト・データを識別結果として出力する。そして図8(c)に示すような状況においては、「これ/が/、/絵本/発明」と「これ/が/、/<エ>/本/発明」という2つの単語列を比較することで、後者がより高い確率であることがより一層、容易に理解される。

【0054】

図8(d)に示したように、冗長語言語モデル322および句読点言語モデル323を用いる場合($\lambda \neq 1$)、話者が、例えば、上述した場合とは異なり、「テン」という音声を省き、音声認識装置1の音声入力装置120(図1)から、「コレガ」・「p a u s e」・「エ」・「ホンハツメイノ」・「p a u s e」・「ヨウテン」という音声を入力すると、音響処理部20は、この音声を示す音素で示す特徴ベクトルを言語復号部22に対して出力する。

【0055】

言語復号部22(図2)は、図7および図12に示したように、音響モデル30、汎用言語モデル320、冗長語言語モデル322および句読点言語モデル323を用いて音響処理部20から入力された特徴ベクトルを処理し、音響モデル30から得られた確率と汎用言語モデル320、冗長語言語モデル322および句読点言語モデル323から得られた確率を評価することにより「コレガ」を「これ/が」と識別し、「これが」の「ガ」に続く「p a u s e」を読点「、」と識別する。つづいて、「エ」「ホン」を「<エ>」「本」と識別する。すなわち「エ」という音に対しては通常語である「絵」などの他に冗長語「<エ>」の可能性も考慮した上で言語モデル32bから得られる確率により、文脈上いずれがより高い可能性を持つかが判定される。そして図8(d)に示すような状況におい

ては、「これ/が/、/絵本/発明」と「これ/が/、/<エ>/本/発明」という2つの単語列を比較することで、後者がより高い確率であることがより一層容易に理解される。さらに、「ハツメイノ」には、「p a u s e」が続くが、「ノ」の後ろには通常、句読点は続かないので、言語復号部22は、この「p a u s e」の部分には句読点を挿入せず、句読点言語モデル323を用いない場合と同様に、入力された音声を「これが、<エ>本発明の要点」と正確に識別し、認識結果のテキスト・データとして出力する。

【0056】

以上の第2の実施形態によれば、第1の実施形態による効果に加えて以下の効果を奏する。すなわち、句読点自動挿入のために、従来からの汎用システムにほとんど変更を加えることなく、句読点自動挿入機能を追加することができる。しかも、ユーザは、この句読点自動挿入用トピックを選択するか否かによって、プログラムの再起動を伴うことなく、句読点自動挿入機能をオン・オフすることができる。

また、文章の内容によって、句読点の挿入頻度を変更したいことがあるが、本実施の形態による句読点挿入方法においては、句読点の出現頻度を、汎用言語モデル320との線形補間時の重みの調整によって容易に制御することができる。

【0057】

以上、2つの実施形態に基づいて本発明を説明してきた。2つの実施形態はいずれも日本語を例にしたものである。冗長語は日本語に限らず英語その他の外国語に適用することもできる。そこで、英語において冗長語が音声認識結果に影響を与える例を示しておく。例文は、She devised remedial measures. である。日本語に翻訳すると、「彼女は善後策を考え出した」となる。ところが、話者が“She”と“devised”との間に<uh>という冗長語を発したとする。すると、<uh>“devised”を“advised”と誤認識するおそれがある。そのような英文について、本実施の形態を適用することにより、冗長語言語モデル322がオフの場合にはShe advised remedial measures. と認識し、冗長語言語モデル322がオンの場合にはShe devised remedial measures. と認識することができる。

【0058】

【発明の効果】

以上説明したように、本発明によれば、汎用的な言語モデルに加えて、冗長語、さらには句読点に特化した語彙・言語モデルを用いることにより、文章中の適切な位置から冗長語等を削除することができ、さらには句読点等の記号を挿入することができる。

【図面の簡単な説明】

【図 1】 第 1 の実施形態による音声認識装置の構成を示す図である。

【図 2】 第 1 の実施形態による音声認識プログラムの構成を示す図である。

【図 3】 図 2 に示した音声認識プログラムの動作を例示する図であって、(a) は汎用言語モデルのみを用いて音声認識が行われる場合場合を示し、(b) は汎用言語モデルおよび冗長語言語モデルを用いて音声認識が行われる場合を示す。

【図 4】 図 2 に示した音声認識プログラム（言語復号部）の処理を示すフローチャートである。

【図 5】 図 2 に示した冗長語言語モデルを用いず、冗長語の自動削除を行わない（トピック・オフの）場合の音声認識プログラム（言語復号部）の動作を示す図である。

【図 6】 図 2 に示した冗長語言語モデルを用い、冗長語の自動削除を行う（トピック・オン）の場合の音声認識プログラム（言語復号部）の動作を示す図である。

【図 7】 第 2 の実施形態による音声認識プログラムの構成を示す図である。

【図 8】 図 7 に示した音声認識プログラムの動作を示す図であって、(a) は汎用言語モデルのみを用いて音声認識が行われる場合を示し、(b) は汎用言語モデルおよび冗長語言語モデル用いて音声認識が行われる場合を、(c) は汎用言語モデルおよび句読点言語モデルを用いて音声認識が行われる場合を、さらに (d) は汎用言語モデル、冗長語言語モデルおよび句読点言語モデルを用いて音声認識が行われる場合を示す。

【図 9】 音声認識プログラム（言語復号部）の処理を示すフローチャートである。

【図 1 0】 図 7 に示した汎用言語モデルのみを用いるため、冗長語の自動削除および句読点の自動挿入を行わない場合の音声認識プログラム（言語復号部）の動作を示す図である。

【図 1 1】 図 7 に示した汎用言語モデルと句読点言語モデルを用いて句読点の自動挿入を行う場合の音声認識プログラム（言語復号部）の動作を示す図である。

【図 1 2】 図 7 に示した汎用言語モデルと冗長語言語モデルを用いて冗長語の自動削除を行う場合の音声認識プログラム（言語復号部）の動作を示す図である。

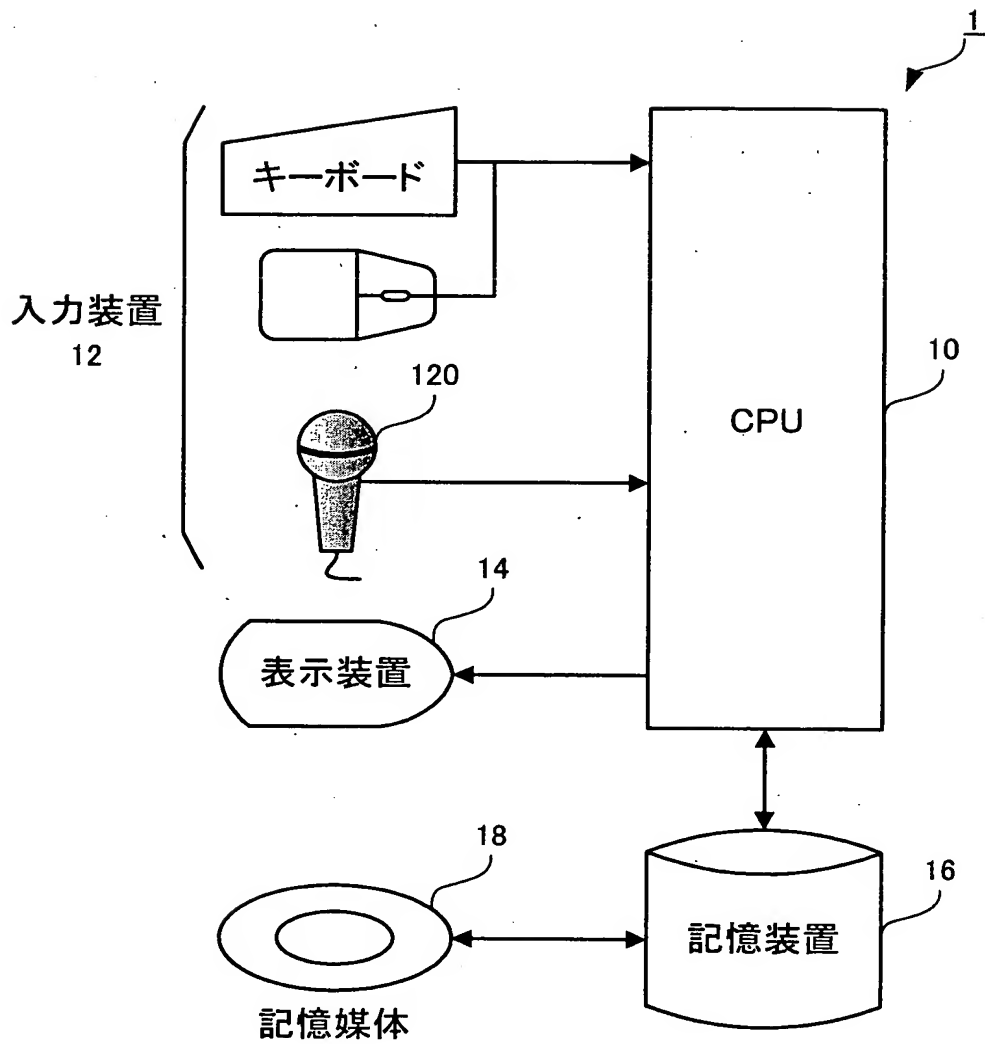
【図 1 3】 図 7 に示した汎用言語モデル、冗長語言語モデルおよび句読点言語モデルを用いて、冗長語の自動削除および句読点の自動挿入を行う場合の音声認識プログラム（言語復号部）の動作を示す図である。

【符号の説明】

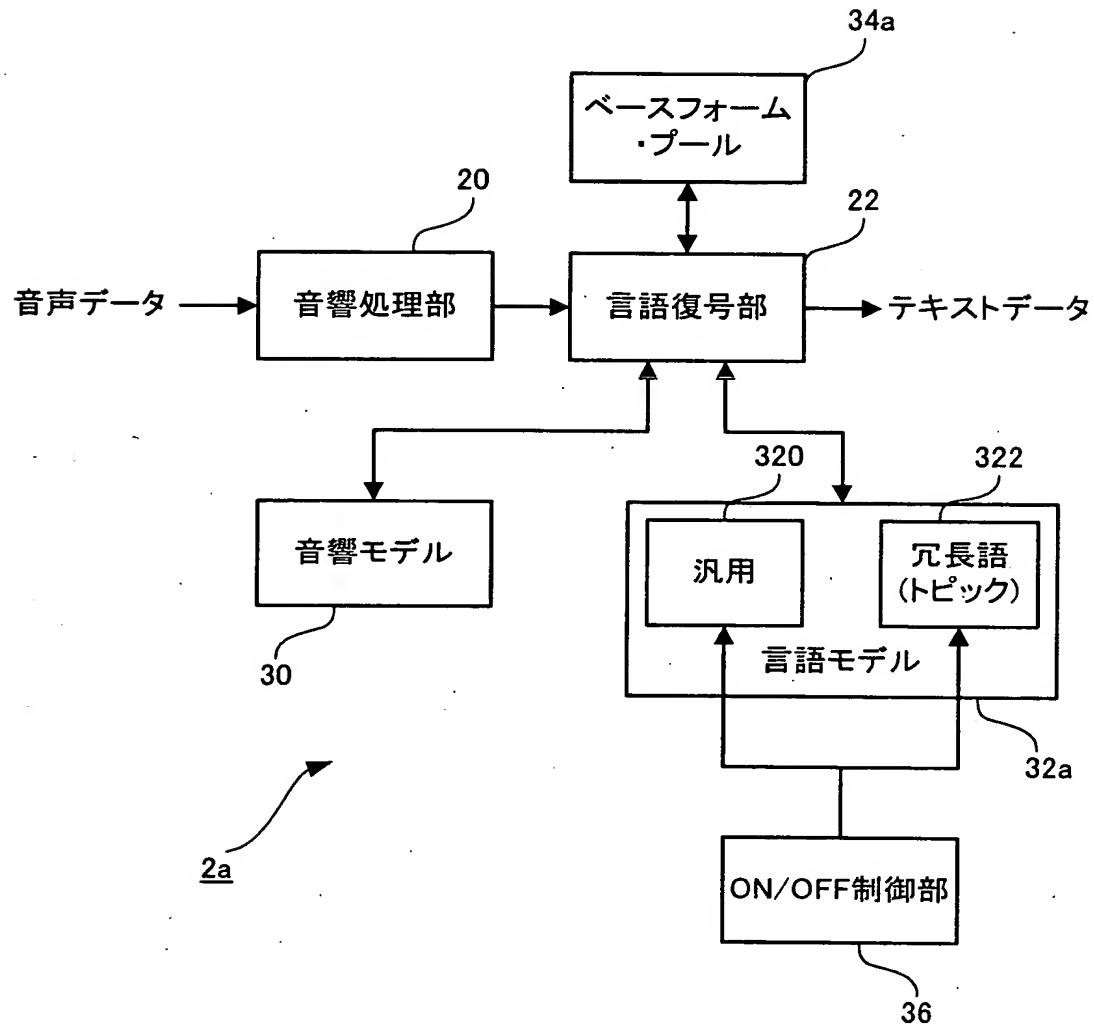
1 … 音声認識装置、1 0 … CPU、1 2 … 入力装置、1 2 0 … 音声入力装置、1 4 … 表示装置、1 6 … 記憶装置、1 8 … 記憶媒体、2 a, 2 b … 音声認識プログラム、2 0 … 音響処理部、2 2 … 言語復号部、3 0 … 音響モデル、3 2 a, 3 2 b … 言語モデル、3 2 0 … 汎用言語モデル、3 2 2 … 冗長語言語モデル、3 2 3 … 句読点言語モデル、3 4 a, 3 4 b … ベースフォーム・プール

【書類名】 図面

【図 1】



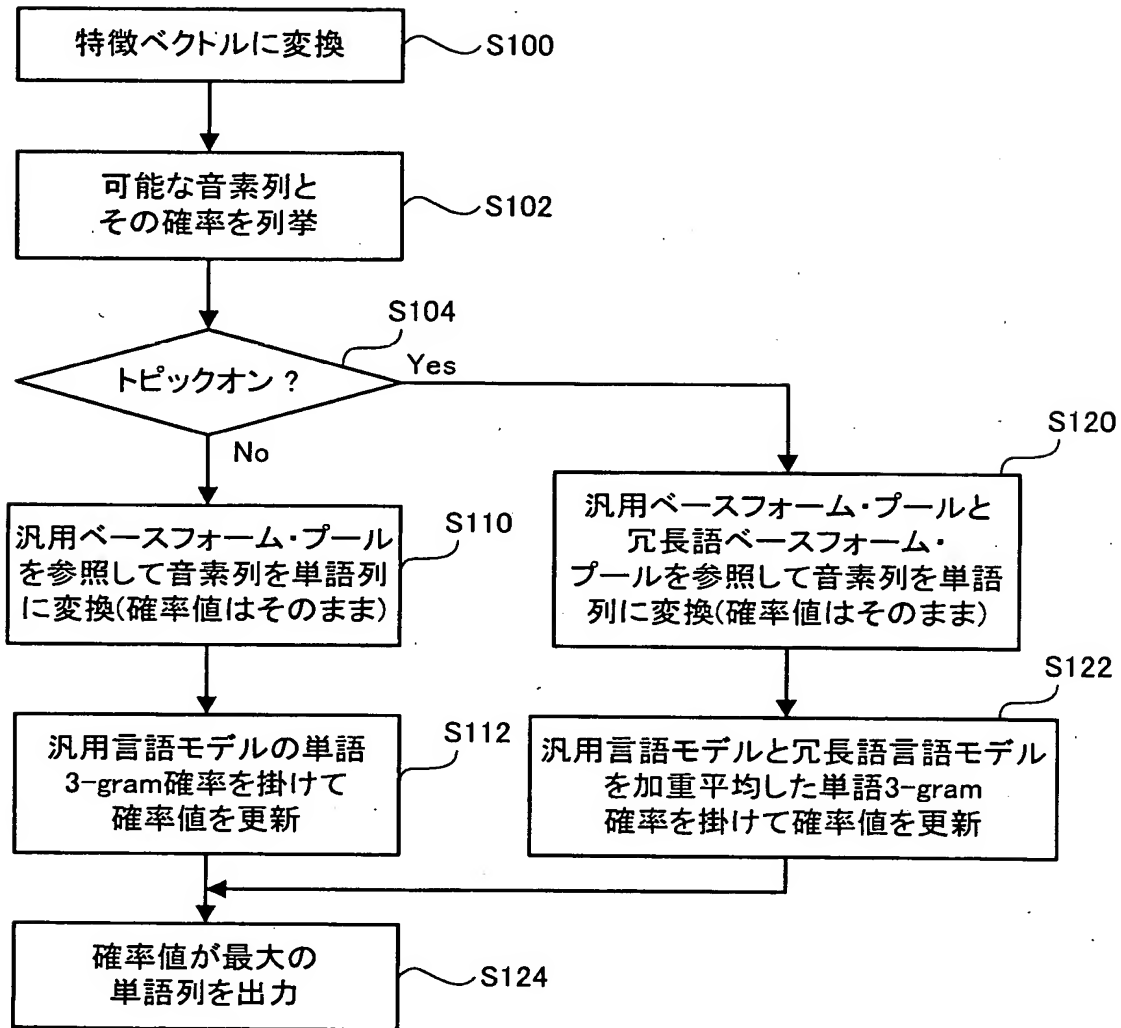
【図 2】



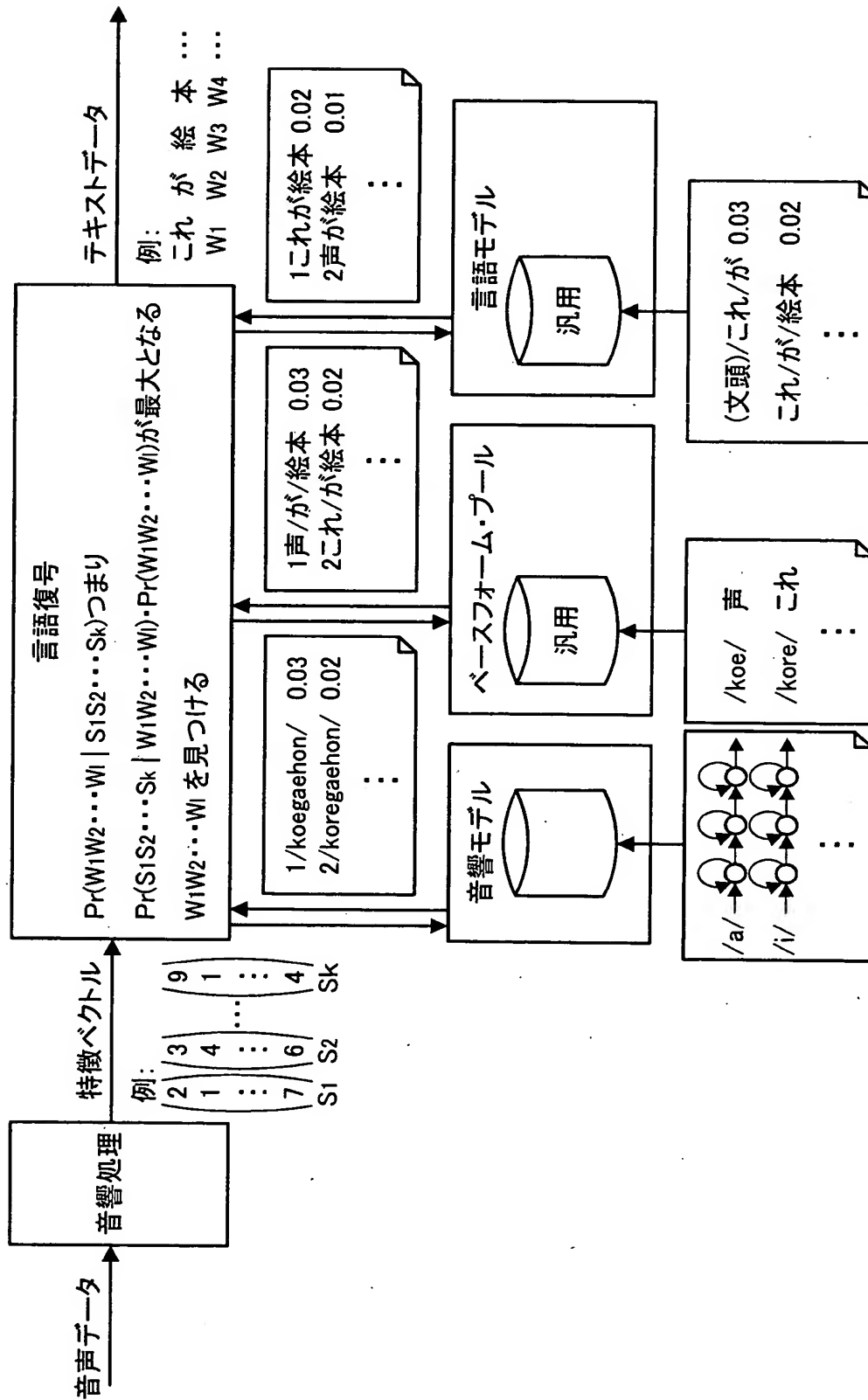
【図 3】

- (a) 冗長語言語モデルがオフの場合
「 ..., コレガエホンハツメイノヨウテン ... 」
→ 「 ..., これが絵本発明の要点 ... 」
- (b) 冗長語言語モデルがオンの場合
「 ..., コレガエホンハツメイノヨウテン ... 」
→ 「 ..., これが<エ>本発明の要点 ... 」

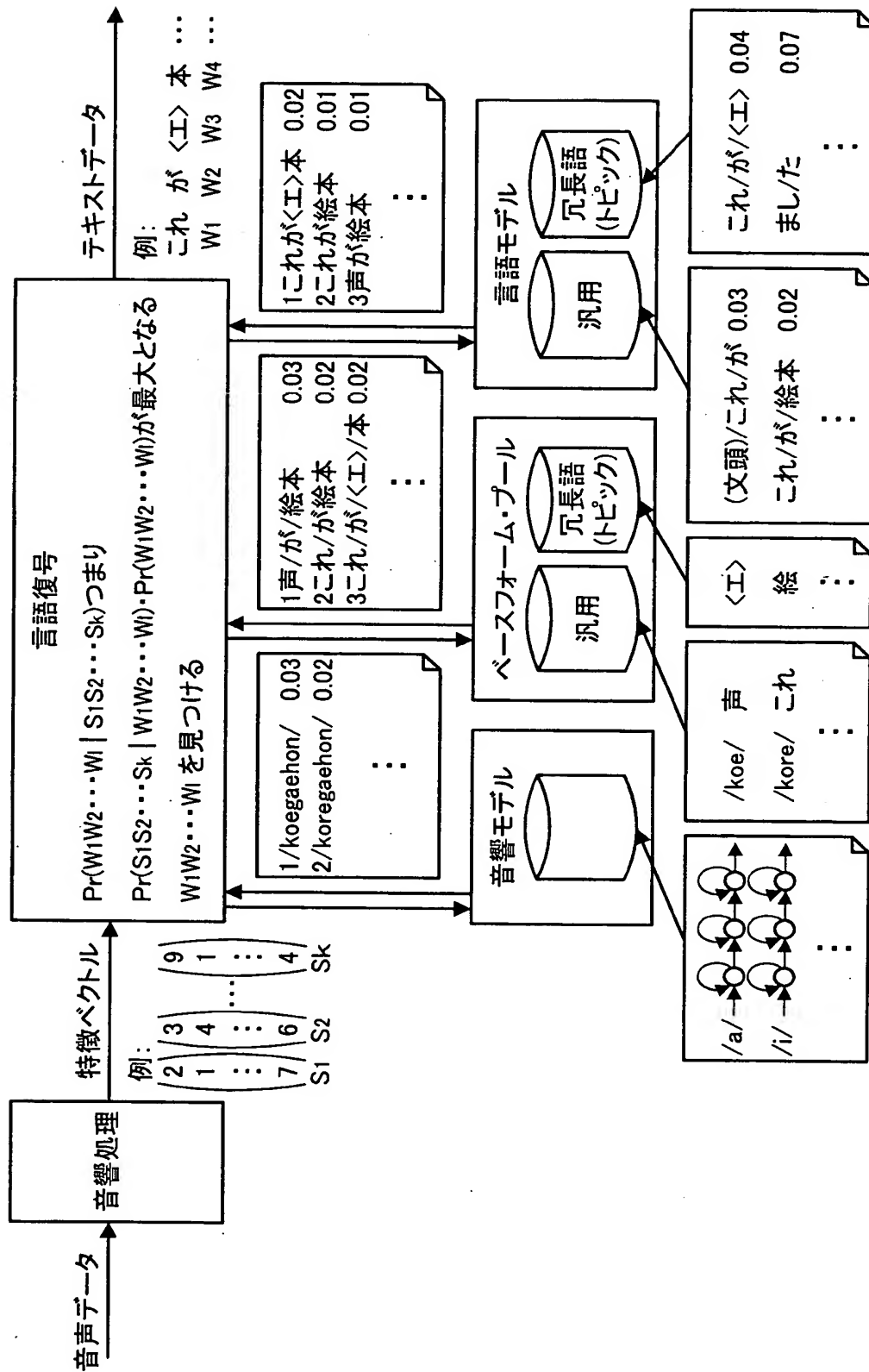
【図 4】



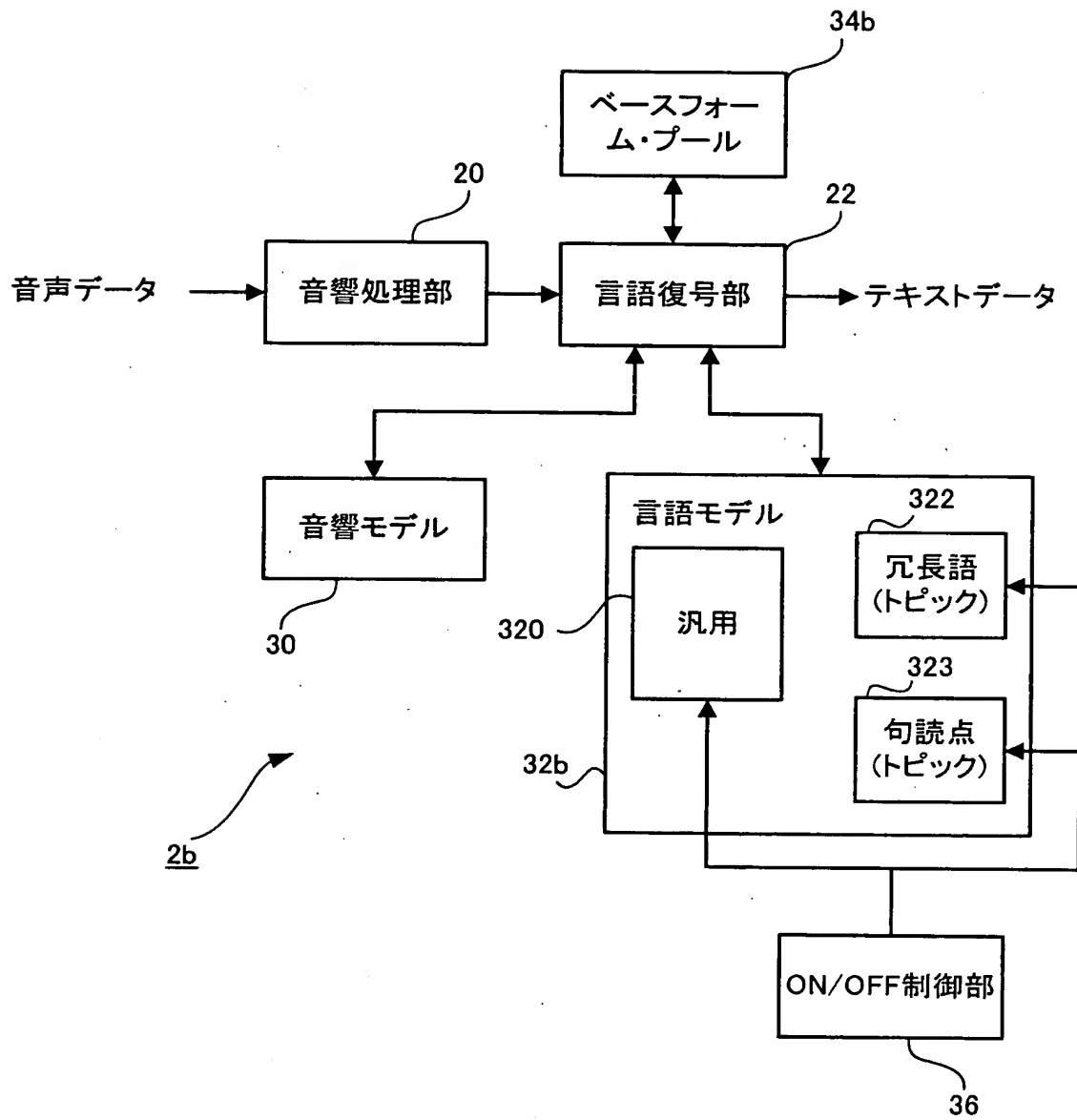
【図 5】



【図 6】



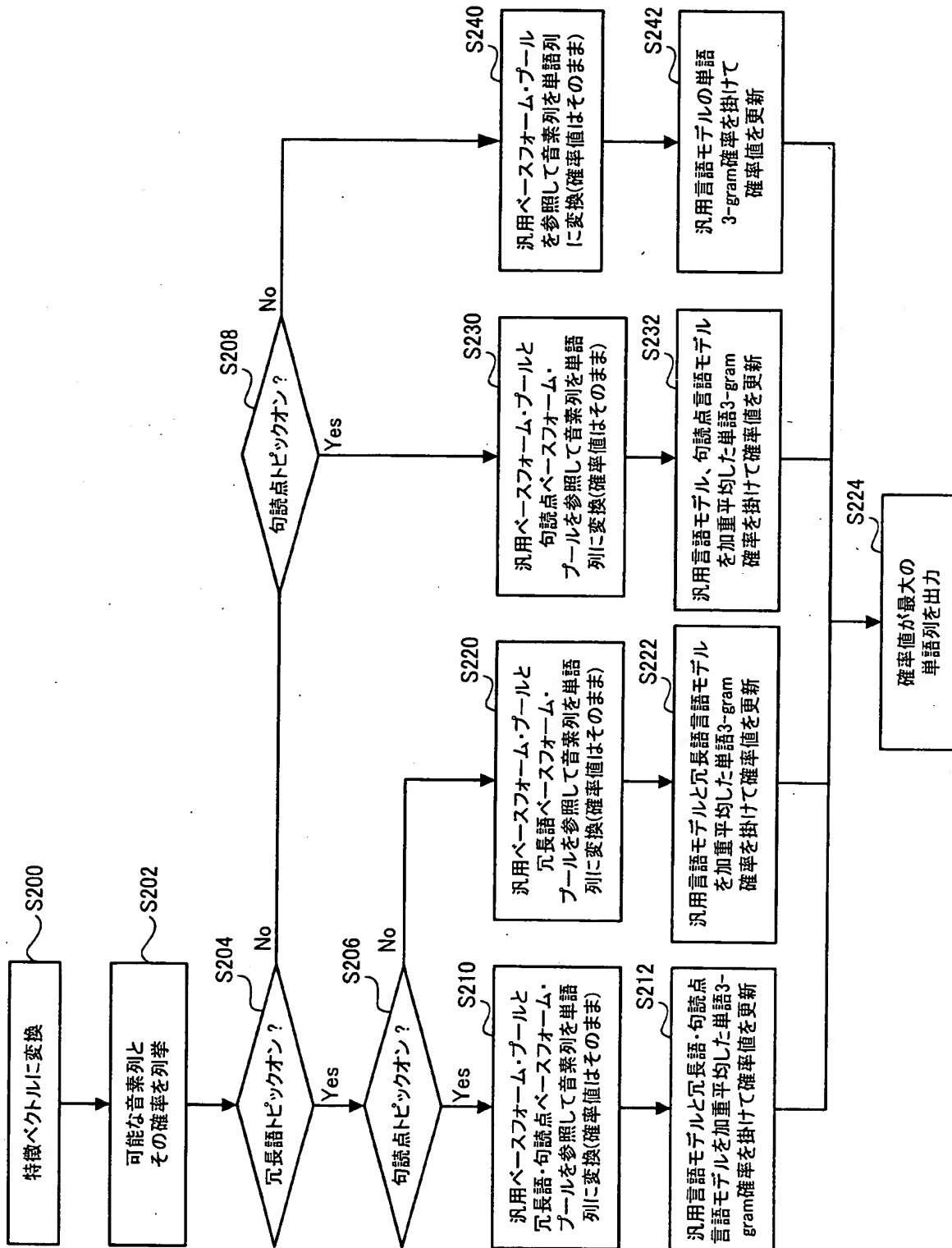
【図 7】



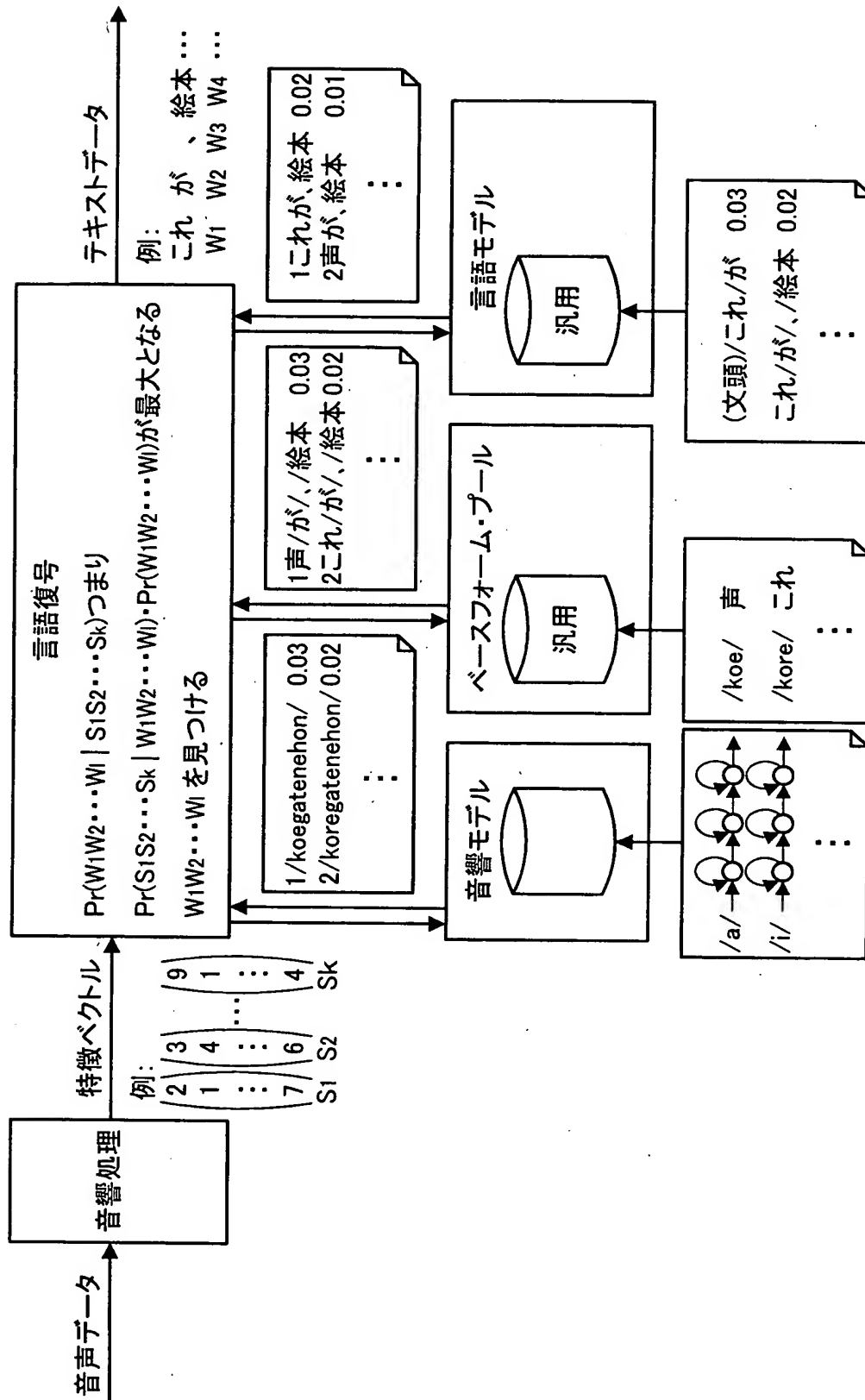
【図 8】

- (a) 冗長言語モデルがオフ、句読点言語モデルがオフの場合
 「 …, コレガ(pause)(テン)エホンハツメイノ(pause)ヨウテン … 」
 ↓
 → 「 …, これが、絵本発明の要点 … 」
- (b) 冗長言語モデルがオフ、句読点言語モデルがオンの場合
 「 …, コレガ(pause)エホンハツメイノ(pause)ヨウテン … 」
 「これ」「が」 — 「、」 — 「絵本」「発明」「の」 — 「、」 — 「要点」
 └─「。」┘ └─「。」┘
 └─NULL ┘ └─NULL ┘
 → 「 …, これが、絵本発明の要点 … 」
- (c) 冗長言語モデルがオン、句読点言語モデルがオフの場合
 「 …, コレガ(pause)(テン)エホンハツメイノ(pause)ヨウテン … 」
 → 「 …, これが、<エ>本発明の要点 … 」
- (d) 冗長言語モデルがオフ、句読点言語モデルがオンの場合
 「 …, コレガ(pause)エホンハツメイノ(pause)ヨウテン … 」
 「これ」「が」 — 「、」 — <エ>「本」「発明」「の」 — 「、」 — 「要点」
 └─「。」┘ └─「。」┘
 └─NULL ┘ └─NULL ┘
 → 「 …, これが、<エ>本発明の要点 … 」

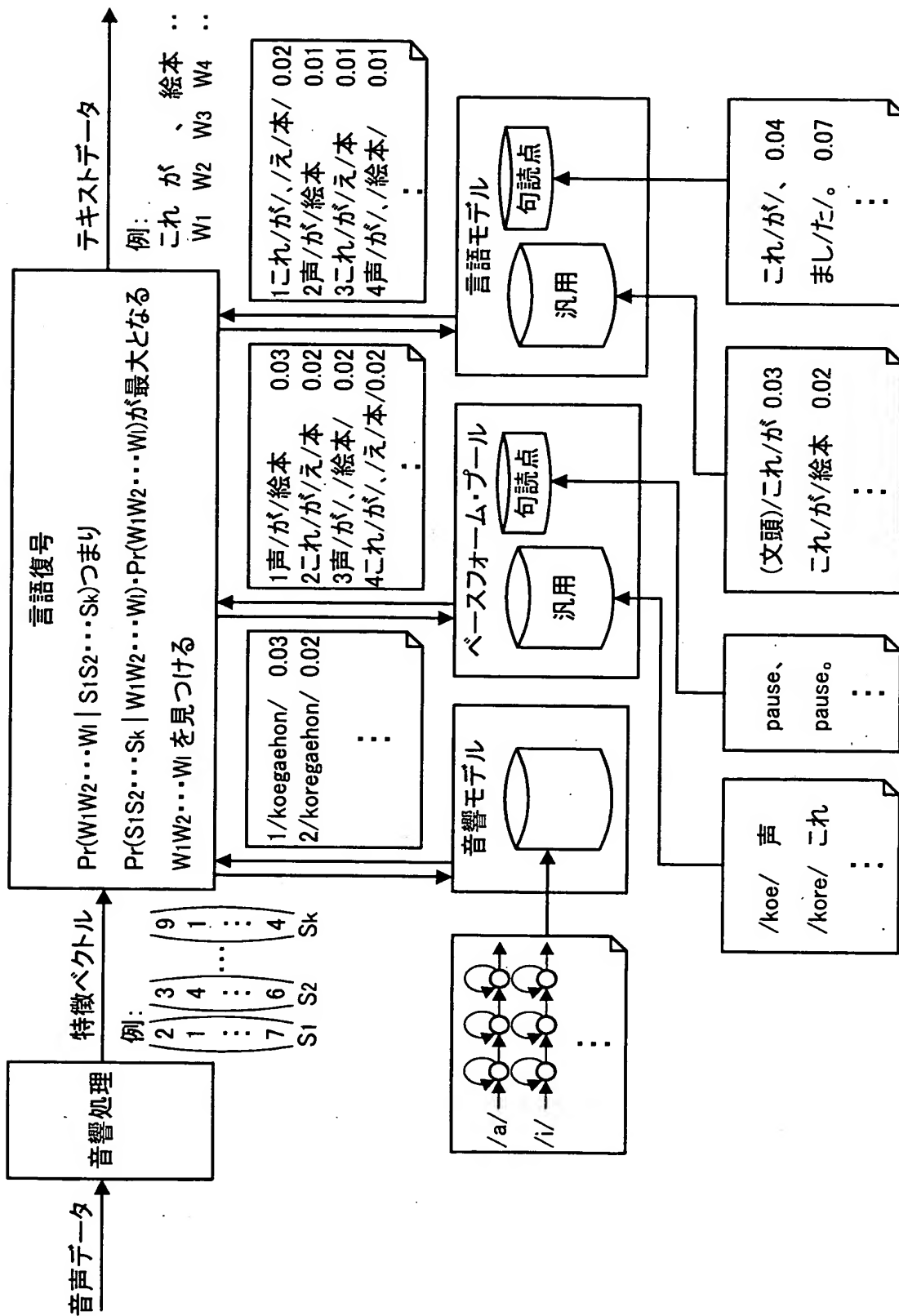
【図 9】



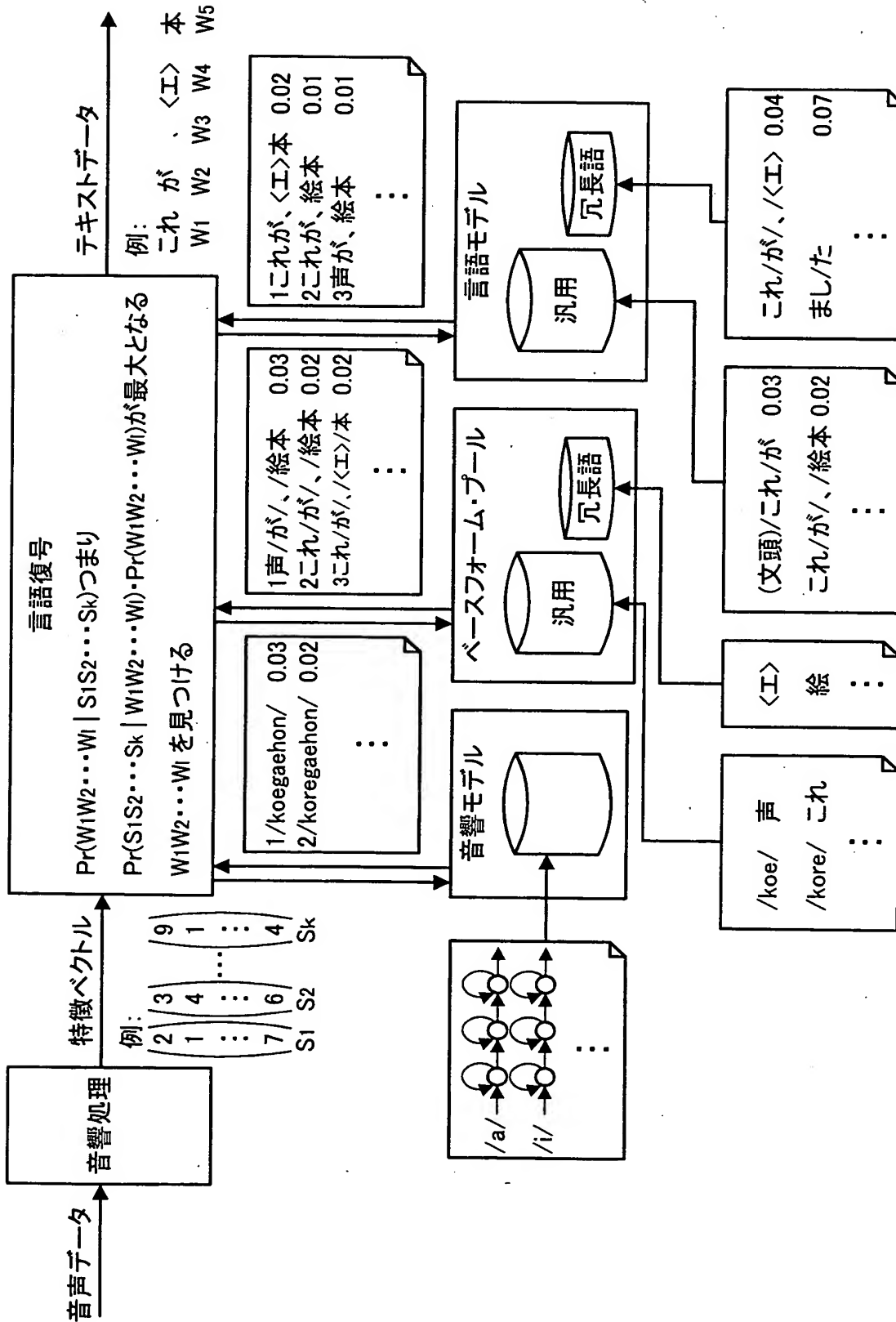
【図10】



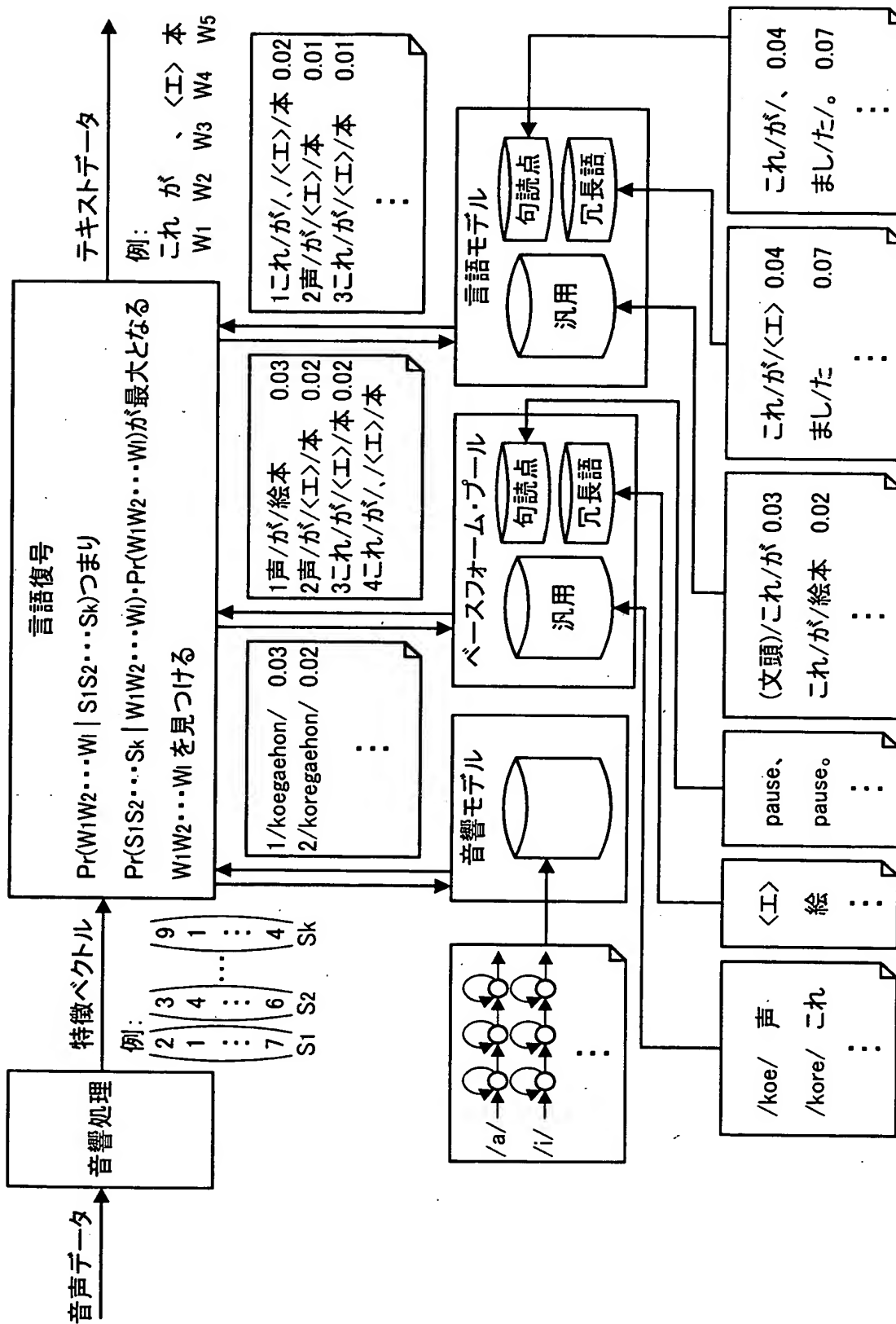
【図 11】



【図 12】



【図 13】



【書類名】 要約書

【要約】

【課題】 ユーザの指定、または話者登録時に得られる情報から、冗長語の除去についてより適しているモデル設定機能を提供する。

【解決手段】 言語モデル 3 2 a として、汎用言語モデル 3 2 0 のほかに冗長語に特化した冗長語言語モデル 3 2 2 を備える。この冗長語言語モデル 3 2 2 を用いることにより、冗長語を自動削除することができる。

【選択図】 図 2

認定・付加情報

特許出願の番号	特願2001-044186
受付番号	50100237522
書類名	特許願
担当官	金井 邦仁 3072
作成日	平成13年 4月 4日

<認定情報・付加情報>

【特許出願人】

【識別番号】	390009531
【住所又は居所】	アメリカ合衆国10504、ニューヨーク州 アーモンク (番地なし)
【氏名又は名称】	インターナショナル・ビジネス・マシーンズ・コーポレーション

【代理人】

【識別番号】	100086243
【住所又は居所】	神奈川県大和市下鶴間1623番地14 日本アイ・ビー・エム株式会社 大和事業所内
【氏名又は名称】	坂口 博

【代理人】

【識別番号】	100091568
【住所又は居所】	神奈川県大和市下鶴間1623番地14 日本アイ・ビー・エム株式会社 大和事業所内
【氏名又は名称】	市位 嘉宏

【代理人】

【識別番号】	100106699
【住所又は居所】	神奈川県大和市下鶴間1623番14 日本アイ・ビー・エム株式会社大和事業所内
【氏名又は名称】	渡部 弘道

【復代理人】

【識別番号】	100104880
【住所又は居所】	東京都港区赤坂5-4-11 山口建設第2ビル 6F セリオ国際特許事務所
【氏名又は名称】	古部 次郎

【選任した復代理人】

【識別番号】	100100077
--------	-----------

次頁有

認定・付加情報（続き）

【住所又は居所】 東京都港区赤坂 5-4-11 山口建設第2ビル
6F セリオ国際特許事務所
【氏名又は名称】 大場 充

出 願 人 履 歴 情 報

識別番号 [390009531]

1. 変更年月日 2000年 5月16日

[変更理由] 名称変更

住 所 アメリカ合衆国10504、ニューヨーク州 アーモンク (番地なし)

氏 名 インターナショナル・ビジネス・マシーンズ・コーポレーション

**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ **BLACK BORDERS**
- ☐ **IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- ☐ **FADED TEXT OR DRAWING**
- ☐ **BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- ☐ **SKEWED/SLANTED IMAGES**
- ☐ **COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- ☐ **GRAY SCALE DOCUMENTS**
- ☐ **LINES OR MARKS ON ORIGINAL DOCUMENT**
- ☐ **REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- ☐ **OTHER:** SMALL Text in Drawings

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.